

**KOMPONEN VARIANS INTERAKSI PELAJAR
DENGAN SOALAN DAN KESAN TERHADAP
KEBOLEHPERCAYAAN**

oleh

TAN YIH TYNG

**Tesis yang diserahkan untuk
memenuhi keperluan bagi
Ijazah Sarjana Sastera**

September 2016

PENGHARGAAN

Penghargaan yang paling tinggi serta ribuan terima kasih yang ditujukan khas untuk penyelia utama saya, Profesor Dr. Abdul Rahman Othman yang telah banyak memberi tunjuk ajar dan galakan kepada saya. Beliau juga menyumbangkan banyak ide yang bernas dalam menyiapkan tesis ini. Beliau juga sudi berjumpa dengan saya pada bila-bila masa walaupun pada hujung minggu atau pada waktu malam. Ribuan terima kasih juga diucapkan kepada penyelia bersama saya, Dr. Teh Sin Yin dan Dr. Lai Choo Heng yang telah memberikan tunjuk ajar serta ide-ide yang berharga dalam penulisan tesis ini.

Penghargaan dan ucapan terima kasih juga ditunjukkan kepada Dr Toh Pek Lan, Cik Chin Ee Laine dan Encik Chiam U-Pin di atas sokongan moral dan bantuan mereka. Selain itu, saya juga mengambil kesempatan untuk mengucapkan terima kasih kepada semua kakitangan akademik dan pentadbiran di Pusat Pengajian Pendidikan Jarak Jauh, USM di atas sokongan dan bantuan mereka pada tempoh sepanjang pengajian saya.

Akhirnya, saya ingin mengucapkan ribuan terima kasih kepada ayah yang dihormati En Tan Gim Bee, ibu yang disayangi Puan Cheah Ah Mooi dan abang yang dikasihi En Tan Yih Shiang. Mereka sering memberikan sokongan moral dan wang di sepanjang tempoh pengajian saya.

JADUAL KANDUNGAN

Penghargaan	ii
Jadual Kandungan	iii
Senarai Jadual	vi
Senarai Rajah	viii
Abstrak	ix
Abstract	xi

BAB SATU – PENGENALAN

1.1	Latar Belakang	1
1.2	Teori Pensampelan	3
1.3	Penyataan Masalah	5
1.4	Objektif Kajian	6
1.5	Sumbangan Kajian	6
1.6	Organisasi Tesis	7

BAB DUA – SOROTAN LITERATUR

2.1	Pengenalan	8
2.2	Teori Ujian Klasik	8
2.2.1	Hubungan di antara Skor Sebenar Dan Skor Tercerap	9
2.2.2	Andaian Ujian Selari	11
2.2.3	Kaedah Untuk Menyediakan Ujian Selari	12
2.3	Teori Kebolehitlakan	15
2.4	Reka Bentuk Faset Tunggal Bersilang	17
2.4.1	Anggaran Komponen Varians dengan ANOVA: Reka Bentuk Faset Tunggal Bersilang	18
2.4.2	Anggaran Komponen Varians dengan ANOVA: Reka Bentuk Dua Faset Bersilang	20
2.5	Pentafsiran Komponen Varians	21
2.5.1	Anggaran Komponen Varians Negatif	23
2.6	Reka Bentuk Tersarang	25
2.7	Anggaran Komponen Varians bagi Model Campuran	28
2.8	Pekali Kebolehitlakan	30

2.9	Interaksi Dalam Teori G	33
 BAB TIGA – METODOLOGI		
3.1	Pengenalan	36
3.2	Pembentukan Ujian	36
3.3	Kajian-kajian G	38
3.4	Pengumpulan Data	40
3.5	Analisis Data	41
3.5.1	Format Data	41
3.5.2	Penghitungan Nilai-nilai Komponen Varians	45
3.5.3	Ujian Tukey Satu-darjah-kebebasan Bagi Sifat Ketidaktambahan Dan Plot Interaksi	51
 BAB EMPAT – HASIL KAJIAN DAN PERBINCANGAN		
4.1	Pengenalan	55
4.2	Hasil Dapatan Kajian G I	55
4.3	Hasil Dapatan Kajian G II	57
4.4	Hasil Dapatan Kajian G III	59
4.5	Hasil Dapatan Kajian G IV	61
4.6	Analisis Ke Atas Tiga Tahap Taksonomi Secara Berasingan	64
4.7	Analisis Tukey Satu-darjah-kebebasan Bagi Sifat Ketidaktambahan	70
4.8	Plot Interaksi	72
 BAB LIMA – KESIMPULAN		
5.1	Pengenalan	77
5.2	Ringkasan Dan Kesimpulan Kajian	77
5.3	Cadangan Untuk Kajian Lanjutan	82
 RUJUKAN		86
LAMPIRAN A		90
LAMPIRAN B		92
LAMPIRAN C		95
LAMPIRAN D		104
LAMPIRAN E		112

LAMPIRAN F	113
LAMPIRAN G	114
SENARAI PENERBITAN	

SENARAI JADUAL

	Muka Surat
Jadual 2.1 Jadual Analisis Varians (ANOVA) Bagi Reka Bentuk Faset Tunggal ($p \times i$)	19
Jadual 2.2 Jangkaan Min Kuasa Dua Bagi Reka Bentuk Dua Faset Bersilang ($p \times i \times o$)	20
Jadual 3.1 Format Data Sebenar Dalam Fail Data Bagi Reka Bentuk Pelajar (p) \times Item (i)	42
Jadual 3.2 Format Data Sebenar Dalam Fail Data Bagi Reka Bentuk Pelajar (p) \times Item (i) \times Masa (o)	43
Jadual 3.3 Format Data Sebenar Dalam Fail Data Bagi Reka Bentuk Pelajar (p) \times Item Tersarang Pada Taksonomi ($i:t$)	44
Jadual 3.4 Format Data Sebenar Dalam Fail Data Bagi Reka Bentuk Pelajar (p) \times item Tersarang Pada Taksonomi ($i:t$) \times Masa (o)	45
Jadual 3.5 Komponen-komponen Varians Yang Dihitung Dalam Keempat-empat Kajian G.	46
Jadual 3.6 Komponen-komponen Varians Yang Perlu Dijangka Dalam Penghitungan Purata Ke Atas Kesan Faset Tetap Bagi Kajian G III Dan Kajian G IV	47
Jadual 3.7 Varians Reja Relatif (σ_{Rel}^2) Bagi Empat-empat Kajian G	47
Jadual 4.1 Analisis Bagi Kajian Pelajar (p) \times Item (i)	56
Jadual 4.2 Analisis Bagi Kajian Pelajar (p) \times Item (i) \times Masa (o)	57
Jadual 4.3 Analisis Bagi Kajian Pelajar (p) \times Item (i) Tersarang Pada Taksonomi ($i:t$)	59
Jadual 4.4 Analisis Bagi Kajian Pelajar (p) \times Item Tersarang Pada Taksonomi ($i:t$) \times Masa (o)	62
Jadual 4.5 Analisis Faset Tetap Secara Berasingan Bagi Kajian Pelajar (p) \times Item Tersarang Pada Taksonomi ($i:t$)	66
Jadual 4.6 Analisis Faset Tetap Secara Berasingan Bagi Kajian Pelajar (p) \times Item Tersarang Pada Taksonomi ($i:t$) \times Masa (o)	67
Jadual 4.7 Analisis Varians Bagi Penilaian Berterusan (PB)	71
Jadual 4.8 Analisis Varians Bagi Penilaian Berterusan (PA)	71
Jadual 4.9 Keupayaan Pelajar Dikelaskan Berdasarkan PNGK	73

Jadual 5.1	Nilai Komponen-komponen Varians Secara Ideal Dan Yang Diperhatikan Di Dalam Kajian G I	81
Jadual 5.2	Nilai Komponen-komponen Varians Secara Ideal Dan Yang Diperhatikan Di Dalam Kajian G II	81

SENARAI RAJAH

		Muka Surat
Rajah 2.1	Gambarajah Venn untuk reka bentuk $p \times (i:t)$	27
Rajah 2.2	Gambarajah Venn yang menunjukkan varians reja relatif bagi reka bentuk faset tunggal bersilang, $p \times i$	32
Rajah2.3	Gambarajah Venn yang menunjukkan varians reja relatif bagi reka bentuk dua faset bersilang, $p \times i \times o$	32
Rajah 3.1	Graf garis yang menggambarkan contoh wujudnya saling tindak di antara keupayaan pelajar dengan item	53
Rajah 3.2	Graf garis yang menggambarkan contoh yang tidak ada saling tindak di antara keupayaan pelajar dengan item	53
Rajah 3.3	Graf garis yang menggambarkan contoh item tidak dapat membezakan jenis keupayaan pelajar	54
Rajah 4.1	Graf garis yang menunjukkan item dengan skor bagi kumpulan keupayaan pelajar dalam data PB	73
Rajah 4.2	Graf garis yang menunjukkan item dengan skor bagi kumpulan keupayaan pelajar dalam data PA	75

KOMPONEN VARIANS INTERAKSI PELAJAR DENGAN SOALAN DAN KESAN TERHADAP KEBOLEHPERCAYAAN

ABSTRAK

Dalam bidang pendidikan, varians interaksi pelajar dengan item merupakan kesan yang terbaur dalam ralat kerana item ujian hanya diuji sekali dan tidak akan diulangi terhadap kumpulan pelajar yang sama. Objektif utama kajian ini adalah untuk mengasingkan kesan interaksi pelajar dengan item daripada unsur ralat. Kesan interaksi pelajar dengan item ini dikaji sama ada ia mempunyai pengaruh terhadap pekali kebolehpercayaan. Teori kebolehitlakan atau dikenali sebagai teori G digunakan dalam kajian ini. Teori G menyediakan dua kajian iaitu kajian kebolehitlakan (kajian G) dan kajian keputusan (kajian D). Kajian G adalah penganggaran komponen varians bagi ujian yang sedia ada kemudian indeks kebolehpercayaan dihitung berdasarkan komponen-komponen varians tersebut. Indeks kebolehpercayaan dalam teori G dikenali sebagai pekali kebolehitlakan, ρ^2 . Pengasingan interaksi pelajar dengan item dari ralat dilakukan berdasarkan kepada empat model yang berasingkan. Model yang pertama (kajian G I) ialah reka bentuk faset tunggal bersilang iaitu pelajar bersilangan dengan item ($p \times i$). Kemudian faset-faset tambahan ditambahkan ke dalam model kedua (kajian G II) iaitu reka bentuk dua faset bersilang pelajar (p) \times item (i) \times masa (o). Model yang ketiga (kajian G III) adalah analogi kepada reka bentuk tersarang bagi model pertama yang mana item tersarang dalam taksonomi (t), reka bentuk $p \times i$ menjadi $p \times i:t$. Hal ini disebabkan oleh tahap taksonomi dalam domain kognitif merupakan kesan tetap dan tahap taksonomi ini adalah rujukan kepada para pendidik semasa penyediaan soalan atau item ujian. Item merupakan kesan rawak dan item ini adalah tersarang di dalam taksonomi. Dalam keadaan seperti ini, kesan taksonomi ke atas pekali kebolehitlakan turut dikaji. Model yang keempat juga adalah analogi kepada reka bentuk tersarang bagi model kedua, yang mana reka bentuk $p \times i \times o$ menjadi $p \times i:t \times o$. Hasil kajian menunjukkan bahawa kesan taksonomi tidak wujud di dalam keempat-empat model tersebut. Hanya model kedua (kajian G II) yang menunjukkan kesan interaksi pelajar dengan item tetapi kesan ini tidak ketara. Ujian Tukey satu-darjah-kebebasan bagi

sifat ketidaktambahan dijalankan ke atas kajian G II bagi tujuan penyemakan keputusan tersebut. Tambahan pula, plot interaksi yang merupakan kaedah penyemakan juga dibina. Kedua-dua ujian penyemakan menunjukkan bahawa kesan interaksi pelajar dengan item adalah wujud tetapi dianggap tidak bererti kerana kesan saling tindaknya adalah lemah.

PERSON BY ITEM INTERACTION VARIANCE COMPONENT AND ITS IMPACT ON RELIABILITY

ABSTRACT

In the field of education, person by item interaction variance component is confounded in error because the test items were only tested once and they will not be repeated to the same group of students. The main objective of this study is to separate the person by item interaction from the error term. The effect of the person by item interaction was studied whether it has an influence on the reliability coefficient. Generalizability theory or also called G theory was used in this study. G theory provides two studies which are generalizability studies (G study) and decision studies (D study). G study is an estimation of variance components for the existing test and the reliability index is calculated based on the variance components. The reliability index in G theory is known as the generalizability coefficient, ρ^2 . Separation of person by item interaction from the error term was done according to four different models. The first model (G study I) is the single facet crossed design which is the person crossed item design ($p \times i$). Then, the supplemental facets were added into the second model (G study II), which are the two facets crossed design, person (p) \times item (i) \times occasion (o). The third model (G study III) was analog to the nested model of the first model, where the item was nested with taxonomy (t) and now $p \times i$ design becomes $p \times i:t$. This is due to the fact that the levels of taxonomy in cognitive domains are the fixed effect and this taxonomy levels were referred by the educators when they are preparing the test questions. The item is a random effect and it is nested in the taxonomy. In this scenario, the impact of taxonomy on reliability coefficient was studied. Similarly, the fourth model (G study IV) was analog to the nested model of the second model, where $p \times i \times o$ design becomes $p \times i:t \times o$. The findings indicate that taxonomy effect does not exist in all the models tested. Only the second model (G study II) shows the effect of the interaction person by item existed, but this effect was not statistically significant. Single-degree-of-freedom for non-

additivity test is run on the second model (G study II) to reexamined the validation. In addition, the interaction plot which also a validation method was constructed. The two validation methods indicate that the interaction person by item has an effect but is considered not significant as the interaction effect is weak.

BAB SATU

Pengenalan

1.1 Latar Belakang

Tujuan utama bagi sesuatu ujian adalah untuk menguji kefahaman dan penguasaan ilmu pengetahuan pelajar terhadap sesuatu topik atau konsep di dalam sesuatu subjek atau mata pelajaran. Setiap topik atau konsep boleh diuji dengan bilangan item yang tak terhingga banyak. Manakala masa yang diperuntukkan untuk setiap ujian adalah terhad. Jadi, bilangan soalan atau item yang disediakan untuk setiap ujian hendaklah berdasarkan masa yang diperuntukkan kepada pelajar untuk mendapatkan penyelesaian.

Walau bagaimanapun, kesemua soalan atau item tersebut tidak dapat digunakan dalam satu ujian. Oleh itu, ujian atau peperiksaan merupakan sampel soalan atau item yang mewakili topik atau konsep (Suen, 1990). Maklum balas pelajar kepada soalan atau item boleh diwakili oleh suatu nombor. Nombor ini dinamakan markah atau skor. Markah atau skor yang diperolehi, iaitu markah yang diberikan oleh pemeriksa dikenali sebagai skor tercerap. Skor ini berkemungkinan mencerminkan atau tidak mencerminkan ciri-ciri pengukuran sebenar bagi ujian tersebut. Berlainan dengan skor tercerap, terdapat skor sebenar yang betul-betul mencerminkan ciri-ciri pengukuran tersebut. Skor sebenar adalah mustahil diperolehi kerana kita tidak mampu menguji semua soalan atau item yang tak terhingga banyak. Jadi, kita mempertimbangkan skor tercerap sebagai skor sebenar.

Kebolehpercayaan boleh ditakrifkan sebagai satu korelasi di antara skor sebenar dan skor tercerap (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991 dan Suen, 1990). Korelasi ini juga dirujuk sebagai indeks kebolehpercayaan. Pekali kebolehitlakan dalam teori kebolehitlakan adalah setara dengan pekali

kebolehpercayaan dalam teori ujian klasik. Pekali kebolehitlakan ini adalah diperoleh daripada penganggaran komponen varians (Brennan, 1992b; Shavelson & Webb, 1991). Komponen-komponen varians tersebut dapat diperoleh daripada reka bentuk dua faktor rawak lengkap yang mengandungi faktor keupayaan pelajar (selepas ini keupayaan pelajar disamaertikan dengan sebutan pelajar), p dan faktor item, i . Berdasarkan reka bentuk ini, kita boleh menganggarkan komponen varians bagi p , i , saling tindak balas p dan i dan ralat. Komponen varians σ_p^2 , σ_i^2 , σ_{pi}^2 dan σ_e^2 masing-masing mewakili pelajar, item, pelajar berinteraksi dengan item dan ralat.

Kita dapat memahami tentang keupayaan pelajar dan darjah perbezaan pengetahuan pelajar daripada nilai komponen varians pelajar, σ_p^2 . Min pula akan memberikan maklumat berkenaan dengan keupayaan pelajar sama ada prestasi kumpulan pelajar ini lemah, sederhana atau cemerlang.

Komponen varians item atau soalan, σ_i^2 , mewakili darjah kesukaran soalan-soalan dalam ujian atau peperiksaan. Di samping itu, kita juga mendapati bahawa σ_i^2 adalah tersarang dalam tahap taksonomi yang terkandung di dalam domain kognitif. Domain kognitif adalah salah satu domain di dalam taksonomi Bloom. Hal ini disebabkan oleh para pendidik yang biasanya menyediakan soalan-soalan ujian adalah berdasarkan domain kognitif ini. Domain kognitif merupakan domain yang melibatkan ilmu pengetahuan dan perkembangan kemahiran intelek (Bloom, 1979). Terdapat enam tahap yang terkandung dalam domain kognitif, mereka adalah pengetahuan (*knowledge*), kefahaman (*comprehension*), aplikasi (*application*), analisis (*analysis*), sintesis (*synthesis*) dan penilaian (*evaluation*) (Bloom, 1979). Enam tahap ini juga merupakan suatu tahap pembelajaran dalam mata pelajaran atau topik tertentu. Di dalam kajian ini, tahap taksonomi tersebut merupakan tahap

kerumitan yang merangkumi bilangan tugas atau langkah penyelesaian yang berbeza-beza.

Komponen varians ralat, σ_e^2 , pula menunjukkan darjah perbezaan yang mungkin dihasilkan oleh perawakan dan daripada sumber-sumber yang tidak diketahui. Sumber-sumber yang tidak diketahui ini seperti kekurangan yakin diri, kelalaian, ketidakcukupan tidur, keletihan dan sebagainya.

Maklumat saling tindak keupayaan pelajar dengan item atau soalan dapat diketahui melalui nilai komponen varians saling tindak pelajar dengan item, σ_{pi}^2 . Komponen varians saling tindak p dengan i , σ_{pi}^2 ini tidak dapat dianggarkan kerana hanya terdapat satu replika di dalam setiap sel saling tindak p dan i . Hal ini disebabkan oleh soalan-soalan ujian biasanya diuji sekali sahaja terhadap kumpulan pelajar yang sama. Jadi kesan saling tindak p dan i adalah terbaaur dalam ralat.

Justeru itu, teori kebolehitlakan yang merupakan salah satu pendekatan bagi teori pensampelan rawak digunakan di dalam kajian ini untuk mengkaji pengasingan varians saling tindak pelajar dengan item, σ_{pi}^2 daripada varians-variens yang lain.

1.2 Teori Pensampelan

Teori pensampelan rawak merupakan salah satu teori psikometri utama yang biasa diguna pakai. Teori pensampelan rawak terdiri daripada dua pendekatan iaitu teori ujian klasik dan teori kebolehitlakan (Crocker & Algina, 1986; Suen, 1990). Teori kebolehitlakan atau dalam Bahasa Inggeris “*Generalizability Theory*” yang merupakan salah satu teori ujian dalam “*Introduction to Classical and Modern Test Theory*” (Crocker & Algina, 1986) digunakan dalam kajian ini. Para pelopor kepada teori ini ialah Cronbach, Gleser dan Rajaratnam (1963). Mereka telah menghuraikan teori tersebut dalam karya mereka yang bertajuk “*Theory of the Generalizability: A Liberalization of Reliability Theory*”.

Teori kebolehitlakan juga dikenali sebagai teori G. Teori G ini menyediakan dua kajian iaitu kajian kebolehitlakan atau dikenali sebagai kajian G dan kajian keputusan yang juga dikenali sebagai kajian D. Kajian kebolehitlakan atau kajian G adalah penganggaran komponen varians bagi ujian yang sedia ada. Pengitlakan ujian tersebut dalam indeks kebolehpercayaan adalah berdasarkan komponen-komponen varians tersebut. Kajian keputusan atau kajian D pula menggunakan maklumat daripada kajian G untuk membentuk suatu pengukuran yang akan meminimumkan reja atau ralat supaya kebolehpercayaan boleh diperbaiki (Shavelson & Webb, 1991; Webb, Shavelson & Haertel, 2006; Shavelson, Webb & Rowley, 1989).

Oleh itu, teori G boleh dikatakan mempunyai hubungan dengan reka bentuk faktorial rawak lengkap dalam bidang Statistik (Brennan, 2001; Shavelson & Webb, 1991). Ia menggunakan komponen-komponen varians daripada reka bentuk tersebut untuk menghitung kebolehpercayaan. Komponen-komponen varians ini juga menyatakan magnitud bagi sumber perbezaan tersebut.

Reka bentuk dalam kajian ini ialah reka bentuk rawak lengkap dua faktor yang melibatkan faktor pelajar, p , iaitu keupayaan pelajar dan faktor item, i , iaitu darjah kesukaran item atau soalan. Faktor p , faktor i dan faktor tindakbalas pi dan ralat masing-masing merupakan sumber perbezaan yang terdapat dalam ujian analisis varians (ANOVA). Dalam teori G, objek pengukuran ialah faktor atau sumber perbezaan dalam ujian ANOVA yang hendak diukur, iaitu faktor p , manakala sumber perbezaan yang lain dipanggil sebagai faset. Dalam perkataan yang lain, faset ini ialah faktor-faktor dalam reka bentuk yang memberikan sumbangan kepada ralat.

Dengan itu, komponen-komponen varians bagi p , i , saling tindak pi dan ralat akan dianggarkan dalam teori G. Indeks kebolehpercayaan dalam kajian G ialah pekali kebolehitlakan, ρ^2 . ρ^2 ialah nisbah varians pelajar dibahagikan oleh jumlah

varians pelajar dan varians ralat. Oleh sebab varians ralat adalah salah satu komponen varians yang memainkan peranan dalam penentuan nilai kebolehpercayaan, jadi, kesan pengasingan komponen varians bagi saling tindak pi daripada varians ralat adalah penting dan amat perlu dikaji dengan lanjut.

1.3 Pernyataan Masalah

Varians saling tindak pelajar dan item, σ_{pi}^2 , biasanya dianggap terbaaur dalam varians ralat, σ_e^2 dan tidak dapat dipisahkan. Maka tertimbul soalan-soalan yang menarik perhatian kita:

- i) Jika soalan-soalan ujian yang sama atau setara diulangi pada kumpulan pelajar yang sama, adakah varians bagi saling tindak pelajar dengan item, σ_{pi}^2 ini terbaaur dalam varians ralat dan tidak ada sebarang kesan?
- ii) Benarkah pengasingan tindak balas p dan i tidak memberi sebarang kesan ke atas varians pelajar, σ_p^2 dan varians item, σ_i^2 ? Justeru itu, pengasingan varians saling tindak pelajar dengan item, σ_{pi}^2 ini perlu diselidiki, lebih-lebih lagi kaitannya dengan varians pelajar, σ_p^2 dan varians item, σ_i^2 . Adakah perkaitan ini berkemungkinan menjejaskan nilai kebolehpercayaan dengan secara tidak langsung?
- iii) Tahap taksonomi domain kognitif digunakan untuk memastikan soalan-soalan yang disediakan dalam ujian ulangan adalah sama atau setara dengan ujian prima (ujian yang diduduki pada kali pertama). Kita mendapati bahawa, varians soalan atau item, σ_i^2 ini adalah tersarang dalam faktor taksonomi. Faktor taksonomi ini merupakan suatu faktor tetap, sedangkan varians soalan atau item pula adalah suatu faktor rawak. Perkara ini telah mewujudkan satu

reka bentuk tersarang. Jadi kesan tetap faktor taksonomi ini perlu dikaji sama ada ia mempunyai pengaruh terhadap kebolehpercayaan ujian.

1.4 Objektif Kajian

Objektif utama kajian ini adalah:

- i) untuk meneliti kesan atau impak yang disumbangkan oleh varians interaksi pelajar dengan item, σ_{pi}^2 ke atas kebolehpercayaan.
- ii) untuk menentukan sama ada terdapat kesan taksonomi (faktor tetap) ke atas kebolehpercayaan.
- iii) untuk memilih model yang terbaik dari aspek parsimoni.

1.5 Sumbangan Kajian

Dengan hasil kajian ini, para pendidik khususnya guru-guru di sekolah akan dapat lebih memahami perbezaan tahap keupayaan pelajar, perbezaan darjah kesukaran item, kesan ralat dan juga kesan saling tindak pelajar dengan item melalui magnitud komponen varians. Kesemua sifat atau kesan disebut di atas amat perlu difahami serta dipertimbangkan oleh para pendidik yang perlu menyediakan soalan ujian.

Dengan wujudnya pengetahuan ini, para pendidik akan dapat menyediakan suatu ujian yang boleh dipercayai, iaitu dengan nilai kebolehpercayaan yang tinggi. Ujian yang boleh dipercayai akan dapat membezakan kumpulan keupayaan pelajar yang lemah, sederhana dan cemerlang.

Dengan kajian ini, para pendidik lebih memahami tentang kesan kesukaran soalan yang tersarang dalam tahap taksonomi dan hubungan tahap taksonomi ini ke atas kesan kesukaran soalan. Tahap taksonomi ini bukan sahaja merupakan tahap pembelajaran, ia juga boleh dijadikan suatu bahan rujukan untuk menyediakan soalan ujian.

1.6 Organisasi Tesis

Di dalam bab ini: latar belakang kajian, teori pensampelan, pernyataan masalah, objektif kajian, sumbangan kajian telah dibentangkan. Organisasi tesis seterusnya adalah seperti berikut: perbincangan tentang konsep-konsep teori pensampelan, reka bentuk ujikaji kajian ini, penganggaran komponen varians serta pentafsiran komponen varians dipersembahkan di dalam BAB DUA. Pembentukan ujian, pengumpulan data dan penganalisan data dibentangkan dalam BAB TIGA. BAB EMPAT adalah perbincangan hasil kajian dan BAB LIMA iaitu bab yang terakhir merupakan rumusan, kesimpulan kajian dan cadangan untuk kajian lanjutan.

BAB DUA

SOROTAN LITERATUR

2.1 Pengenalan

Huraian di dalam bab ini adalah berkenaan dengan teori psikometri yang digunakan dalam kajian ini, iaitu teori kebolehitlakan. Oleh sebab teori ujian klasik merupakan induk kepada teori kebolehitlakan, jadi konsep-konsep yang terkandung dalam teori ujian klasik dibincangkan terdahulu dan diikuti dengan konsep-konsep yang terkandung dalam teori kebolehitlakan. Perbincangan yang seterusnya ditumpukan pada reka bentuk kajian ini serta penganggaran komponen varians dan pentafsirannya bagi reka bentuk kajian ini. Kemudian pekali kebolehitlakan dihuraikan. Kajian-kajian yang berkenaan dengan interaksi dibincangkan di akhir bab ini.

2.2 Teori Ujian Klasik

Apabila kita menyebut “ujian”, satu gambaran yang muncul dalam pemikiran kita ialah peperiksaan atau ujian dalam bidang akademik. Hal ini disebabkan ujian merupakan satu prosedur yang penting. Ujian mesti dialami oleh semua pelajar sama ada dalam pendidikan prima, sekunder atau pendidikan tinggi. Ujian adalah sampel item yang mewakili topik atau konsep sesuatu subjek, kerana kita tidak mampu menguji semua soalan atau item di dalam satu ujian. Skor yang diperoleh oleh pelajar daripada ujian dirujuk sebagai maklum balas pelajar terhadap item. Skor yang diperoleh daripada ujian digelar sebagai skor tercerap. Semua sifat atau ciri sebenar bagi ujian berkemungkinan tidak dapat dicerminkan oleh skor tercerap. Skor yang betul-betul mencerminkan ciri-ciri sebenar ialah skor sebenar. Skor sebenar ini adalah mustahil diperoleh, kerana kita tidak dapat menguji item yang tak terhingga

banyak di dalam sesuatu ujian. Justeru itu, skor tercerap yang dipertimbangkan oleh penyelidik bidang psikometrik.

Teori ujian klasik merupakan teori pengukuran yang paling awal. Kini, teori ujian klasik ini masih mempunyai pengaruh yang kuat di kalangan ujian pengukuran yang diamalkan. Teori ujian klasik juga dirujuk sebagai teori kebolehpercayaan klasik kerana ia mengutamakan penganggaran kebolehpercayaan skor tercerap (Crocker & Algina, 1986; Lei, Smith & Suen, 2007; Suen, 1990).

Kebolehpercayaan boleh ditakrifkan sebagai satu korelasi di antara skor sebenar dan skor tercerap. Korelasi ini dirujuk sebagai indeks kebolehpercayaan dan juga dinyatakan sebagai korelasi Pearson. Pekalinya ditandakan dengan ρ_{xt} yang mana x ialah skor tercerap dan t ialah skor sebenar (Crocker & Algina, 1986; Suen, 1990). Apabila nilai korelasi semakin tinggi, maka bertambah baik skor tercerap mencerminkan skor sebenar. Malangnya, ρ_{xt} tidak dapat dianggar secara terus daripada data tercerap kerana dalam dunia nyata skor sebenar, t adalah mustahil diketahui. Walau bagaimanapun, kita dapat menganggar nilai kuasa dua korelasi Pearson, ρ_{xt}^2 .

2.2.1 Hubungan Di Antara Skor Sebenar Dan Skor Tercerap

Skor sebenar dirujuk sebagai min bagi skor yang diperolehi daripada semua ujian yang diuji dengan tak terhingga banyak kali. Skor tercerap pula dirujuk sebagai skor yang diperolehi dalam sesuatu ujian. Nilainya berkemungkinan lebih atau kurang daripada skor sebenar. Hal ini disebabkan terdapat banyak sumber yang tidak diketahui seperti keletihan, kelalaian dan lain-lain yang menjejaskan skor tercerap. Perbezaan di antara skor tercerap dan skor sebenar dikenali sebagai skor ralat.

Secara matematik, hubungan skor tercerap dengan skor sebenar bagi sesuatu ujian tunggal dapat dinyatakan sebagai:

$$x = t + e \quad (2.1)$$

yang mana x ialah skor tercerap, t ialah skor sebenar dan e ialah skor ralat (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991; Suen, 1990). Korelasi Pearson ρ di antara dua pembolehubah sebarangan X dan Y adalah dipaparkan dalam Persamaan (2.2) (Crocker & Algina, 1986; Larsen & Marx, 2006; Suen, 1990):

$$\rho_{xy} = \frac{\text{cov}(X,Y)}{(\text{sisihan piawai } X)(\text{sisihan piawai } Y)} \quad (2.2)$$

yang mana $\text{cov}(X, Y)$ ialah kovarians bagi X dan Y . Maka korelasi di antara skor tercerap dan skor sebenar, ρ_{xt} dapat dituliskan seperti berikut:

$$\rho_{xt} = \frac{\text{cov}(x,t)}{\sqrt{\text{var}(x)\text{var}(t)}} \quad (2.3)$$

yang mana $\text{cov}(x,t)$ ialah kovarians bagi x dan t , $\text{var}(x)$ ialah varians bagi x atau varians skor tercerap dan $\text{var}(t)$ ialah varians bagi t atau varians skor sebenar. Persamaan (2.4) diperolehi dengan menggantikan model skor sebenar, $x = t + e$ ke dalam Persamaan (2.3)

$$\rho_{xt} = \frac{\text{var}(t) + \text{cov}(t,e)}{\sqrt{\text{var}(x)\text{var}(t)}} \quad (2.4)$$

yang mana $\text{cov}(t,e)$ ialah kovarians bagi t dan e .

Skor sebenar dianggap tidak ada hubungan dengan skor ralat. Justeru itu t dan e tidak bersandar, maka $\text{cov}(t,e)$ bersamaan dengan sifar, kemudian kuasadrakan Persamaan (2.4), Persamaan yang baru dapat ditulis sebagai:

$$\rho_{xt}^2 = \frac{\text{var}(t)}{\text{var}(x)} \quad (2.5)$$

Menurut Crocker dan Algina (1986), Pedhazur dan Schmelin (1991) dan Suen (1990), Persamaan (2.5) ialah varians skor sebenar dibahagi dengan varians skor tercerap. Persamaan ini juga dikenali sebagai pekali kebolehppercayaan. Kita dapat menganggarkan pekali kebolehppercayaan, ρ_{xt}^2 dengan andaian ujian selari.

2.2.2 Andaian Ujian Selari

Ujian selari merupakan dua ujian (ujian A dan ujian B) yang sengaja direka setara untuk mengukur keupayaan pelajar dalam kumpulan yang sama. Andaikan hubungan skor tercerap dengan skor sebenar bagi ujian A adalah

$$X_A = T + E_A$$

yang mana X_A ialah skor tercerap ujian A, T ialah skor sebenar dan E_A ialah reja atau ralat ujian A.

Andaian hubungan skor tercerap dengan skor sebenar bagi ujian B adalah

$$X_B = T + E_B$$

yang mana X_B ialah skor tercerap ujian B, T ialah skor sebenar dan E_B ialah reja atau ralat ujian B.

Korelasi bagi skor tercerap, indeks kebolehppercayaan di antara kedua-dua set ujian A dan ujian B adalah:

$$\rho_{AB} = \frac{\text{var}(T) + \text{cov}(E_{X_A} E_{X_B})}{\sqrt{\text{var}(X_A) \text{var}(X_B)}}. \quad (2.6)$$

Dalam andaian ujian selari, terdapat dua anggapan khusus iaitu:-

- (i) Skor ujian A dan ujian B mempunyai nilai varians yang sama,

$$\text{var}(X_A) = \text{var}(X_B) = \text{var}(X).$$

- (ii) Ralat bagi ujian A dan ujian B adalah saling tak bersandar, oleh yang demikian $\text{cov}(E_{X_A} E_{X_B}) = 0$.

Dengan anggapan-anggapan khusus tersebut dan kuasaduakan Persamaan (2.6), pekali kebolehpercayaan adalah ditunjukkan seperti berikut:

$$\rho_{AB}^2 = \frac{\text{var}(T)}{\text{var}(X)} = \rho_{xt}^2. \quad (2.7)$$

Pekali kebolehpercayaan ini sebenarnya ialah nilai Pearson ρ dikuasaduakan. Oleh itu, selang nilai pekali kebolehpercayaan adalah dari 0 hingga 1 sahaja dan kebolehpercayaan yang bernilai negatif adalah melanggar andaian ujian selari. Jika andaian ujian selari tidak dipenuhi, korelasi hanya menggambarkan hubungan di antara kedua-dua skor ujian tersebut dan tidak mencerminkan hubungan di antara skor tercerap dan skor sebenar. Jadi kepastian ujian-ujian memenuhi andaian ujian selari adalah penting.

2.2.3 Kaedah Untuk Menyediakan Ujian Selari

Kaedah yang paling mudah untuk menyediakan dua ujian selari ialah penilaian berulangan. Dalam kaedah ini, satu ujian yang sama akan diulangi pada masa yang berlainan ke atas kumpulan pelajar yang sama. Selang masa di antara dua ujian tersebut mungkin menyebabkan skor sebenar berubah. Selepas ujian pertama, pelajar-pelajar mungkin membincangkan jawapan atau mencari penyelesaian yang betul. Mereka akan menambahbaikkkan prestasi mereka pada ujian kedua iaitu ujian ulangan dengan tindakan ini. Ini akan menjejaskan nilai kebolehpercayaan. Justeru itu, pekali ρ_{xt}^2 dalam penilaian berulangan adalah suatu nilai anggaran.

Kaedah yang lebih berkesan adalah mereka bentuk satu ujian dengan memecahkan soalan-soalan dalam ujian kepada dua bahagian yang sama. Kedua-dua bahagian dianggap selari iaitu mengandungi bilangan dan tahap kesukaran soalan yang sama. Kaedah ini juga dikenali sebagai pecah-separuh (*split-half*) (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991; Suen, 1990). Pekali

kebolehpercayaan yang diperolehi daripada kaedah ini merupakan nilai kebolehpercayaan bagi separuh bahagian sahaja. Jadi, anggaran pekali kebolehpercayaan ini adalah tidak begitu tepat, kerana bilangan item dalam ujian telah dikurangkan daripada bilangan item dalam ujian yang sebenar. Malahan, nilai ralat juga akan turut menurun kerana kandungan dalam sampel telah berkurangan. Biasanya bilangan item ujian yang lebih akan memberikan nilai pekali kebolehpercayaan yang lebih tinggi berbanding daripada bilangan item ujian yang kurang (Crocker & Algina, 1986). Masalah ini dapat diperbaiki dengan menggunakan kaedah Spearman-Brown.

Menurut Crocker dan Algina (1986); Suen (1990), rumus Spearman-Brown merupakan rumus yang umum untuk menganggarkan kebolehpercayaan, ρ_{xt}^2 apabila bilangan item dalam ujian ditambahkan ataupun dikurangkan. Persamaan (2.8) memaparkan rumus Spearman-Brown, yang mana N^* adalah bilangan item baru yang diunjurkan dan bukan bilangan item yang asal. Kebolehpercayaan bagi jumlah skor ujian dengan bilangan item yang diunjurkan, N^* ditanda dengan:

$$\rho_{xt}^{2*} = \frac{N^* \rho_{xt}^2}{1 + (N^* - 1) \rho_{xt}^2} \quad (2.8)$$

di mana ρ_{xt}^2 ialah kebolehpercayaan ujian asal. Menurut Brown (2002), kebolehpercayaan ini juga dikenali sebagai kebolehpercayaan konsistensi dalaman.

Terdapat pelbagai cara untuk memecahkan ujian kepada beberapa ujian separuh. Setiap ujian pecahan separuh akan memberikan anggaran kebolehpercayaan yang berbeza. Ini merupakan satu kelemahan kebolehpercayaan yang diperolehi daripada kaedah pecah separuh.

Kelemahan tersebut boleh diatasi dengan indeks “Cronbach’s alpha”. Rumus pekali ini dipaparkan dalam Persamaan (2.9)

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right) \quad (2.9)$$

yang mana σ_i^2 ialah varians bagi setiap item ke- i , σ_x^2 adalah jumlah varians skor cerapan dan N adalah jumlah item dalam ujian. Kebolehpercayaan konsistensi dalaman yang diperolehi dengan rumus ini merupakan batas bawah kepada pelbagai kebolehpercayaan yang diperolehi daripada Persamaan (2.8) (Brown, 2002).

Menurut Brown (2002), pekali kebolehpercayaan konsistensi dalaman ialah pekali Kuder-Richardson atau KR-20. Rumus KR-20 dipaparkan dalam Persamaan (2.10):

$$KR-20 = \frac{N}{N-1} \left(1 - \frac{\sum p_i q_i}{\sigma_x^2} \right). \quad (2.10)$$

Sebenarnya, rumus KR-20 dan rumus “Cronbach’s alpha” adalah sama. Pembolehubah $p_i q_i$ dalam Persamaan (2.10) adalah rumus bagi varians σ_i^2 dalam taburan Bernoulli. Persamaan p_i mewakili kadar pelajar yang menjawab dengan betul, manakala q_i adalah kadar pelajar yang menjawab dengan salah untuk item i . σ_x^2 adalah jumlah varians skor cerapan dan N adalah jumlah item dalam ujian. KR-20 boleh dinyatakan sebagai kes yang khusus bagi “Cronbach’s alpha” bagi semua item berbentuk dikotomi.

Di dalam teori ujian klasik, ralat tidak dipertimbangkan dan dianggap memperoleh nilai yang sama di dalam semua keadaan ujian. Ralat ini tidak digambarkan dengan jelas dan dikenali sebagai ralat rawak (Suen, 1990). Ralat ini juga dianggap sebagai suatu pembolehubah rawak yang tidak dibezakan (Webb, Shavelson & Haertel, 2006). Sebenarnya, ralat perlu dipertimbangkan kerana

nilainya berkemungkinan berbeza pada setiap ujian. Ralat ini merangkumi sumber-sumber yang tidak diketahui seperti ketinggalan menjawab soalan, kerosakan alat-alat tulis dan sebagainya. Sumber-sumber tersebut tidak akan berlaku pada setiap ujian, justeru itu nilai ralat akan berbeza. Selain itu, teori ujian klasik juga tidak boleh mempunyai lebih daripada satu pekali kebolehppercayaan dalam set data yang sama, kerana pekali kebolehppercayaan yang dihitung berdasarkan Pearson ρ adalah pekali kebolehppercayaan yang tunggal.

Teori kebolehitlakan yang diperkenalkan oleh Cronbach et al. (1963) merupakan perkembangan teori ujian klasik yang mana sumber-sumber ralat dipertimbangkan dan diperolehi.

2.3 Teori Kebolehitlakan

Teori ujian klasik dan teori kebolehitlakan merupakan dua pendekatan yang terdapat dalam teori pensampelan yang biasa diguna pakai (Suen, 1990). Teori kebolehitlakan atau dikenali sebagai teori G adalah satu teori statistik yang menilai kebolehppercayaan pengukuran tingkah laku (Brennan, 2001; Cronbach, et al., 1972; Shavelson & Webb, 1991). Menurut Brennan (1992a, 2001), teori kebolehitlakan juga merupakan satu teori yang menggabungkan sifat-sifat teori ujian klasik dan analisis varians (ANOVA). Oleh itu, kita boleh mengatakan bahawa teori ujian klasik dan ANOVA ialah induk kepada teori kebolehitlakan.

Apabila suatu pengukuran diulangi, kita berkemungkinan besar akan memperoleh nilai-nilai yang berbeza. Perbezaan ini mungkin disebabkan oleh perubahan pada instrumen pengukuran, perubahan pada keadaan pengukuran atau perubahan kepada ciri pengukuran yang sedang diukur. Kita boleh menyimpulkan bahawa perbezaan nilai cerapan ini disebabkan oleh kewujudan varians sistematik. Menurut teori ujian klasik, sesuatu skor itu terdiri daripada skor sebenar dan ralat.

Ralat ini tidak mencerminkan perbezaan yang mungkin disebabkan oleh perubahan-perubahan yang dinyatakan di atas. Oleh yang demikian, varians sistematik tidak akan dipertimbangkan oleh teori ujian klasik.

Teori kebolehitlakan atau teori G boleh mengasingkan ralat di dalam teori ujian klasik kepada komponen-komponen yang sistematik dan rawak. Teori G juga boleh mengoptimumkan kebolehppercayaan instrumen pengukuran. Di dalam kajian kebolehitlakan atau kajian G, sumber-sumber perbezaan yang boleh menjurus kepada perbezaan skor di antara pelajar perlu dikenal pasti. Seterusnya, magnitud sumber-sumber perbezaan dianggar dengan menggunakan kaedah komponen varians di dalam kajian ANOVA. Dengan anggaran magnitud sumber-sumber perbezaan, pekali kebolehitlakan dapat dihitung.

Pekali kebolehitlakan yang dihitung mungkin tidak mencapai nilai kebolehppercayaan yang tinggi. Nilai kebolehppercayaan 0.80 ke atas dianggap sebagai tinggi (Salvia & Ysseldyke, 1991). Jika nilai kebolehppercayaan kurang daripada 0.80, kita boleh mengkaji pertambahan kepada nilai tersebut dengan cara-cara berikut:

1. Kekalkan semua komponen varians yang diperolehi daripada sumber perbezaan sistematik. Kemudian, laraskan perhitungan kebolehppercayaan dengan memasukkan bilangan tahap sumber perbezaan yang baru sehingga nilai kebolehppercayaan mencapai 0.80 atau lebih.
2. Kekalkan bilangan tahap sumber perbezaan, tingkatkan atau kurangkan komponen varians, melebihi atau mengurangi magnitud asal sehingga kebolehppercayaan bersamaan dengan 0.80 atau lebih. Peningkatan atau pengurangan magnitud baru ini mestilah berdasarkan kemungkinan komponen varians yang munasabah.

3. Berdasarkan kemungkinan yang munasabah, komponen varians dan bilangan tahap sumber perbezaan ditingkatkan atau dikurangkan supaya nilai kebolehpercayaan mencapai 0.80 atau lebih.

Kajian pengoptimuman ini dikenali sebagai kajian keputusan, di dalam Bahasa Inggeris dikenali sebagai “*Decision Study*” atau kajian D.

2.4 Reka Bentuk Faset Tunggal Bersilang

Pertimbangkan reka bentuk faktorial yang melibatkan dua faktor: A dan B . Persamaan model tersebut diberikan seperti berikut:

$$Y_{ij} = \mu + A_i + B_j + \varepsilon_{ij}, i = 1, 2, \dots, a; j = 1, 2, \dots, b. \quad (2.11)$$

yang mana Y_{ij} ialah skor item ke- j yang dicapai oleh pelajar ke- i . Pekali μ ialah min keseluruhan skor. Pembolehubah A_i ialah kesan pelajar ke- i , B_j ialah kesan item ke- j dan ε_{ij} ialah ralat. Jika faktor A dan B merupakan faktor rawak, maka $A_i \sim N(0, \sigma_A^2)$, $B_j \sim N(0, \sigma_B^2)$ dan $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

Di dalam teori kebolehitlakan reka bentuk ini dikenali sebagai reka bentuk faset tunggal. Faktor A ialah objek pengukuran. Objek pengukuran ialah objek pengitlakan ujian. Di dalam contoh ini, ujian diitlakkan ke atas pelajar. Dalam senario pendidikan, ujian juga boleh diitlakkan ke atas kelas, sekolah mahupun negeri. Ini dapat diperhatikan dalam surat khabar apabila keputusan atau gred-gred SPM dan STPM diumumkan setiap tahun. Faktor yang selebihnya, yakni B dikenali sebagai faset, sesuatu faset itu adalah sumber kepada ralat pengukuran. Faktor B merupakan sumber perbezaan yang sistematik. Kesukaran soalan dapat membezakan pencapaian pelajar (objek pengukuran). Sebarang faktor yang menyumbangkan perbezaan sistematik selain daripada item dikenali sebagai faset. Untuk reka bentuk ini, penilai juga boleh menjadi faset selain daripada item. Sebagai contohnya, di

dalam penilaian kemahiran masakan seperti *Master Chef Malaysia*, 20 orang peserta dinilai oleh 3 orang penilai.

2.4.1 Anggaran Komponen Varians dengan ANOVA: Reka Bentuk Faset Tunggal Bersilang

Teori G menggunakan ANOVA untuk memisahkan skor individu kepada kesan bagi objek pengukuran, kesan bagi setiap faset dan kesan bagi setiap sumber interaksi. Dalam reka bentuk faset tunggal bersilang yang asal, pembolehubah A bersilang dengan pembolehubah B ($A \times B$), sekarang ditanda dengan $p \times i$ iaitu pelajar bersilang dengan item. Sumber variasi skor ialah kesan pelajar (p), kesan soalan atau item (i), kesan ralat yang mengandungi interaksi pelajar dengan soalan dan pelbagai perubahan yang tidak dapat dicerap (pi,e). Oleh sebab setiap pelajar dinilai hanya sekali pada setiap item, maka ralat adalah gabungan interaksi pelajar dengan item dan reja ukuran. Kadangkala kesan ralat ini juga dikenali sebagai kesan interaksi.

Daripada ANOVA, reka bentuk pelajar bersilang soalan, hasil tambah kuasa dua jumlah (SS_t) dihasilkan daripada

$$SS_t = SS_p + SS_i + SS_{pi,e} \quad (2.12)$$

yang mana SS_p = hasil tambah kuasa dua jumlah kesan pelajar, SS_i = hasil tambah kuasa dua jumlah kesan soalan atau item, $SS_{pi,e}$ = hasil tambah kuasa dua jumlah kesan ralat. Min kuasa dua (MS) bagi setiap kesan ialah hasil tambah kuasa dua dibahagikan oleh darjah kebebasan kesan tersebut. Jumlah pelajar ialah n_p dan jumlah item ialah n_i . Jadual 2.1 memaparkan maklumat yang diuraikan tadi.

Jadual 2.1

Jadual Analisis Varians (ANOVA) Bagi Reka Bentuk Faset Tunggal ($p \times i$)

Sumber Variasi	Hasil Tambah Kuasa Dua (SS)	Darjah Kebebasan (df)	Min Kuasa Dua (MS)
Pelajar (p)	SS_p	$n_p - 1$	$MS_p = \frac{SS_p}{n_p - 1}$
Item (i)	SS_i	$n_i - 1$	$MS_i = \frac{SS_i}{n_i - 1}$
Ralat (pi,e)	$SS_{pi,e}$	$(n_p - 1)(n_i - 1)$	$MS_{pi,e} = \frac{SS_{pi,e}}{(n_p - 1)(n_i - 1)}$

Tatatanda n_p ialah jumlah pelajar, n_i ialah jumlah item.

Jangkaan MS_p , MS_i dan $MS_{pi,e}$, diperolehi melalui

$$E(MS_p) = \sigma_{pi,e}^2 + n_i \sigma_p^2, \quad (2.13)$$

$$E(MS_i) = \sigma_{pi,e}^2 + n_p \sigma_i^2 \text{ dan} \quad (2.14)$$

$$E(MS_{pi,e}) = \sigma_{pi,e}^2. \quad (2.15)$$

Anggaran-anggaran bagi komponen varians di dalam Persamaan (2.13), (2.14) dan (2.15) dapat ditulis sebagai:

$$\hat{\sigma}_{pi,e}^2 + n_i \hat{\sigma}_p^2 = MS_p, \quad (2.16)$$

$$\hat{\sigma}_{pi,e}^2 + n_p \hat{\sigma}_i^2 = MS_i \text{ dan} \quad (2.17)$$

$$\hat{\sigma}_{pi,e}^2 = MS_{pi,e}. \quad (2.18)$$

Dengan menyusun semula Persamaan (2.16) dan Persamaan (2.17) terhadap $\hat{\sigma}_p^2$ dan

$\hat{\sigma}_i^2$ masing-masing, kita perolehi

$$\hat{\sigma}_p^2 = \frac{MS_p - \hat{\sigma}_{pi,e}^2}{n_i} \text{ dan} \quad (2.19)$$

$$\hat{\sigma}_i^2 = \frac{MS_i - \hat{\sigma}_{pi,e}^2}{n_p}. \quad (2.20)$$

Langkah-langkah penganggaran komponen varians ini dapat dilanjutkan kepada reka bentuk pelbagai faset (Brennan, 2001; Burdick & Graybill, 1992; Shavelson & Webb, 1991).

2.4.2 Anggaran Komponen Varians dengan ANOVA: Reka Bentuk Dua Faset Bersilang

Pertimbangkan reka bentuk dua faset bersilang dan faktor-faktor yang terlibat ialah faktor pelajar (p), faktor soalan atau item (i) dan faktor masa (o). Ketiga-tiga faktor tersebut dianggap sebagai sampel rawak. Set soalan kursus Teori Kebarangkalian diulangi pada masa yang berlainan ke atas pelajar-pelajar tahun tiga di PPPJJ, USM. Maka, reka bentuk ialah pelajar bersilang dengan item dan masa, ditandakan sebagai $p \times i \times o$. Reka bentuk ini mengandungi tujuh sumber variasi dan juga tujuh komponen varians sepadan yang dapat dianggarkan serta ditafsirkan. Langkah-langkah untuk penganggaran komponen varians akan dibincangkan pada Bahagian 2.5 dalam bab ini. Jadual 2.2 memaparkan rumus bagi jangkaan min kuasa dua bagi reka bentuk dua faset bersilang.

Jadual 2.2

Jangkaan Min Kuasa Dua Bagi Reka Bentuk Dua Faset Bersilang ($p \times i \times o$)

Sumber Variasi	Komponen Varians	Jangkaan Min Kuasa Dua (EMS)
pelajar (p)	σ_p^2	$\sigma_{pio,e}^2 + n_o \sigma_{pi}^2 + n_i \sigma_{po}^2 + n_i n_o \sigma_p^2$
item (i)	σ_i^2	$\sigma_{pio,e}^2 + n_p \sigma_{io}^2 + n_o \sigma_{pi}^2 + n_p n_o \sigma_i^2$
masa (o)	σ_o^2	$\sigma_{pio,e}^2 + n_p \sigma_{io}^2 + n_i \sigma_{po}^2 + n_p n_i \sigma_o^2$
pelajar x item (pi)	σ_{pi}^2	$\sigma_{pio,e}^2 + n_o \sigma_{pi}^2$
pelajar x masa (po)	σ_{po}^2	$\sigma_{pio,e}^2 + n_i \sigma_{po}^2$
item x masa (io)	σ_{io}^2	$\sigma_{pio,e}^2 + n_p \sigma_{io}^2$
ralat (pio,e)	$\sigma_{pio,e}^2$	$\sigma_{pio,e}^2$

Komponen-komponen varians dianggarkan seperti berikut:

$$\hat{\sigma}_p^2 = \frac{MS_p - \hat{\sigma}_{pio,e}^2 - n_o \hat{\sigma}_{pi}^2 - n_i \hat{\sigma}_{po}^2}{n_i n_o}, \quad (2.21)$$

$$\hat{\sigma}_i^2 = \frac{MS_i - \hat{\sigma}_{pio,e}^2 - n_p \hat{\sigma}_{io}^2 - n_o \hat{\sigma}_{pi}^2}{n_p n_o}, \quad (2.22)$$

$$\hat{\sigma}_o^2 = \frac{MS_o - \hat{\sigma}_{pio,e}^2 - n_p \hat{\sigma}_{io}^2 - n_i \hat{\sigma}_{po}^2}{n_p n_i}, \quad (2.23)$$

$$\hat{\sigma}_{pi}^2 = \frac{MS_{pi} - \hat{\sigma}_{pio,e}^2}{n_o}, \quad (2.24)$$

$$\hat{\sigma}_{po}^2 = \frac{MS_{po} - \hat{\sigma}_{pio,e}^2}{n_i}, \quad (2.25)$$

$$\hat{\sigma}_{io}^2 = \frac{MS_{io} - \hat{\sigma}_{pio,e}^2}{n_p}, \text{ dan} \quad (2.26)$$

$$\hat{\sigma}_{pio,e}^2 = MS_{pio,e}. \quad (2.27)$$

Tatatanda n_p ialah bilangan pelajar, n_i ialah bilangan soalan dan n_o ialah masa bilangan selang. Varians interaksi pelajar×item×masa terbaaur dalam varians ralat.

2.5 Pentafsiran Komponen Varians

Pentafsiran komponen varians dalam reka bentuk faset tunggal bersilang dibincangkan dahulu, kemudian diikuti dengan pentafsiran komponen varians bagi reka bentuk dua faset bersilang.

Komponen varians untuk kesan pelajar, σ_p^2 mencerminkan darjah perbezaan pengetahuan di antara seorang pelajar dengan pelajar lain dari segi prestasi pencapaian atau darjah perbezaan pencapaian pelajar-pelajar tersebut. Jika skor tertabur rapat di sekitar sesuatu min sut kesan pelajar, hanya segelintir pelajar yang berskor tinggi dan rendah, maka skor ini akan ditunjukkan dengan nilai anggaran relatif yang rendah bagi komponen varians pelajar. Manakala, nilai relatif bagi

komponen varians pelajar yang besar pula menunjukkan pelajar mempunyai sebaran pencapaian yang luas. Sebagai heuristik dalam pentafsiran magnitud relatif anggaran komponen varians, kita boleh mengambil jumlah komponen varians kemudian menghitungkan peratusan bagi setiap varians.

Komponen varians bagi kesan soalan atau item, σ_i^2 menunjukkan darjah perbezaan di antara soalan atau item dari segi kesukaran. Serupa dengan komponen varians pelajar, komponen varians item mencerminkan kesukaran soalan atau item berbanding min sut kesan soalan dan nilainya juga boleh menjadi besar atau kecil relatif kepada komponen-komponen varians pelajar dan ralat.

Komponen varians untuk ralat, $\sigma_{pi,e}^2$, menunjukkan darjah perbezaan yang dihasilkan oleh perawakan dan daripada sumber-sumber yang tidak diketahui. Kesan interaksi pelajar dengan item, $p \times i$ dalam reka bentuk faset tunggal juga bercampur aduk dalam komponen varians ralat. Saiz varians interaksi secara relatif dengan saiz sumber-sumber variasi yang sistematik tetapi dikenal pasti ataupun tidak sistematik diketahui. Contoh sumber variasi sistematik yang tidak dikenal pasti adalah perbezaan keadaan ujian (masa dan suasana). Manakala, sumber tidak sistematik ialah perlakuan yang tidak sengaja seperti tertinggal soalan, kelalaian, kerosakan alat tulis atau alat penghitung semasa menduduki peperiksaan.

Dalam reka bentuk dua faset bersilang, σ_p^2 dan σ_i^2 mempunyai tafsiran yang bersamaan dengan komponen varians yang terkandung di dalam reka bentuk faset tunggal bersilang. Komponen varians untuk kesan masa, σ_o^2 yang besar menunjukkan bahawa masa memberi kesan kepada prestasi pelajar. Sebagai contoh, pelajar-pelajar berkemungkinan dapat skor yang lebih pada masa ujian yang berlainan.

Komponen varians kesan interaksi pelajar dengan item, σ_{pi}^2 bernilai besar menunjukkan bahawa kedudukan relatif kelompok pelajar berbeza daripada satu kelompok soalan ke satu kelompok soalan. Komponen varians kesan interaksi pelajar dengan masa, σ_{po}^2 pula bermakna kedudukan relatif kelompok prestasi pelajar berbeza daripada satu kelompok masa ke satu kelompok masa. Komponen varians interaksi item dengan masa, σ_{io}^2 , menggambarkan kesukaran item berbeza daripada satu kelompok masa ke satu kelompok masa yang lain.

Akhirnya, kesan interaksi pelajar \times item \times masa, σ_{pio}^2 terbaaur di dalam komponen varians ralat, $\sigma_{pio,e}^2$. Seperti reka bentuk faset tunggal bersilang, komponen varians ini juga mengandungi sumber-sumber sistematik tetapi tidak dapat dikenal pasti atau sumber-sumber tidak sistematik.

2.5.1 Anggaran Komponen Varians Negatif

Kadang kala, anggaran komponen varians dalam sesuatu sampel mungkin negatif. Hal ini adalah mustahil dari segi teori nilai anggaran negatif ini adalah mustahil. Justeru itu, Shavelson dan Webb (1991) dan Briesch et al.(2014) telah mencadangkan penggunaan satu pendekatan yang dikemukakan oleh Cronbach et al. (1972) dalam penerbitan mereka untuk menyelesaikan masalah ini, iaitu menentukan anggaran varians yang bernilai negatif ini menjadi sifar. Terdapat dua kaedah yang mengaplikasikan pendekatan ini. Cara yang pertama melibatkan pengiraan komponen varians di dalam sistem persamaan. Jika beberapa komponen varians yang diperolehi pada peringkat awal adalah negatif, maka nilai negatif tersebut akan digantikan dengan sifar. Di dalam penyelesaian komponen varians yang lain pada peringkat kedua yang memerlukan nilai komponen-komponen varians peringkat pertama, nilai-nilai sifar dikenakan. Selain itu, Shavelson dan Webb (1991) juga

mencadangkan cara lain untuk mengaplikasikan pendekatan ini. Kesemua anggaran komponen varians tidak kira sama ada negatif ataupun positif dikumpulkan. Kemudian, hanya gantikan kesemua komponen varians yang negatif dengan nilai sifar.

Satu lagi pendekatan untuk mengelakkan anggaran varians menjadi negatif adalah menggunakan kaedah kebolehjadian maksimum (Webb, Shavelson & Haertel, 2006).

Katakan y_1, y_2, \dots, y_n adalah sampel rawak daripada taburan yang mempunyai fungsi ketumpatan sebarangan dengan parameter, θ . Maka, fungsi kebolehjadian y_1, y_2, \dots, y_n dapat ditulis

$$\begin{aligned} L(\theta; y_1, y_2, \dots, y_n) &= L(\theta), \\ &= \prod_{i=1}^n f(y_i; \theta) \end{aligned} \quad (2.28)$$

Penganggar kebolehjadian maksimum θ dan penyelesaian persamaan adalah ditunjukkan seperti berikut (Ahmad, 1996):

$$\frac{d}{d\theta} L(\theta) = 0. \quad (2.29)$$

Pertimbangan reka bentuk faset tunggal bersilang yang dipaparkan dalam Persamaan (2.11), parameter-parameter yang terkandung di dalam Persamaan (2.11) adalah $\mu, \sigma_A^2, \sigma_B^2$ dan σ_ε^2 . Maka fungsi kebolehjadian dapat ditulis

$$L(\mu, \sigma_A^2, \sigma_B^2, \sigma_\varepsilon^2 | y_{11}, y_{12}, \dots, y_{ij}). \quad (2.30)$$

Jadi penganggaran maksimum bagi $\mu, \sigma_\tau^2, \sigma_\beta^2$ dan σ_ε^2 dapat diperoleh seperti persamaan-persamaan yang dipaparkan:

$$\frac{\partial}{\partial \mu} L(\mu, \sigma_A^2, \sigma_B^2, \sigma_\varepsilon^2) = 0 \quad (2.31)$$